

①9 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENTAMT

⑫ Offenlegungsschrift
⑪ DE 3436818 A1

②① Aktenzeichen: P 34 36 818.3
②② Anmeldetag: 6. 10. 84
②③ Offenlegungstag: 10. 4. 86

⑤① Int. Cl. 4:
C 12 N 15/00
C 12 N 1/20
C 12 R 1/19
C 12 P 19/34
C 07 H 21/04

Behördenentwurf

DE 3436818 A1

⑦① Anmelder:
Hoechst AG, 6230 Frankfurt, DE

⑦② Erfinder:
Engels, Joachim, Dr., 6242 Kronberg, DE;
Leineweber, Michael, Dr., 6230 Frankfurt, DE;
Uhlmann, Eugen, Dr., 6240 Königstein, DE;
Wetekam, Waldemar, Dr., 6239 Eppstein, DE

⑤④ Synthetische Signalsequenz zum Transport von Proteinen in Expressionssystemen

Die DNA einer natürlichen Signalsequenz wird durch Einbau von Schnittstellen für Endonucleasen modifiziert und lässt sich dadurch nach dem Baukastenprinzip in beliebige Vektoren einbauen. Die derart modifizierten Vektoren bewirken dann einen Transport des codierten Proteins aus dem Zytoplasma.

DE 3436818 A1

Patentansprüche:

HOE 84/F

1. Synthetische Signalsequenz zum Transport von Proteinen in Expressionssystemen, dadurch gekennzeichnet, daß die DNA im wesentlichen einer natürlichen Signalsequenz entspricht, aber eine oder mehrere Schnittstellen für
5 Endonucleasen aufweist, die in der natürlichen DNA nicht enthalten sind.
2. Signalsequenz nach Anspruch 1, dadurch gekennzeichnet, daß sie interne Schnittstellen vor und/oder nach der
10 hydrophoben Region enthält.
3. Signalsequenz nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß sie im wesentlichen der natürlichen Signalsequenz der Alkalischen Phosphatase von E. coli
15 entspricht.
4. Signalsequenz nach einem oder mehreren der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß sie am
20 3'-Ende bis zu etwa 40 der aminoterminalen Codons des Strukturgens enthält.
5. DNA der Formel I.
6. Verfahren zur Transportexpression von Proteinen in
25 prokaryotischen und eukaryotischen Zellen, dadurch gekennzeichnet, daß man das Gen für das zu transportierende Protein an eine DNA-Sequenz nach Anspruch 1 bis 5 koppelt, dieses Fusionsgen in einen Vektor einbaut und damit eine Wirtszelle transformiert,
30 die das exprimierte Protein aus dem Cytoplasma transportiert.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet,
daß das transportierte Protein ein evkaryotisches,
prokaryotisches oder virales Protein ist.
- 5 8. Verfahren nach Anspruch 6 oder 7, dadurch gekennzeichnet,
daß die synthetische DNA-Signalsequenz ein wirts-
eigenes Protein kodiert.
9. Hybridvektor, gekennzeichnet durch eine DNA-Sequenz
nach Anspruch 1 bis 5.
- 10 10. Hybridplasmid, das in einer Eco RI-Schnittstelle die
DNA-Sequenz I inseriert enthält.
11. Wirtsorganismus, enthaltend einen Vektor nach Anspruch
15 9 oder 10.
12. E. coli, enthaltend einen Vektor nach Anspruch 9 oder
10.

3.

Synthetische Signalsequenz zum Transport von Proteinen in
Expressionssystemen

Proteine werden in der Zelle an den Ribosomen synthetisiert, die im Cytoplasma lokalisiert sind. Proteine, die aus dem Cytoplasma heraus transportiert werden, tragen am Amino-terminus eine relativ kurze Peptidkette, die beim Passieren
5 durch die Cytoplasmamembran enzymatisch abgetrennt wird, wobei das "reife" Protein entsteht. Diese kurze Peptidsequenz wird als "Signalpeptid" oder auch als Prae- bzw. leader-Sequenz bezeichnet.

- 10 Es wurde schon für eine Vielzahl sekretorischer Proteine die am Aminoterminus angeordnete Signalsequenz charakterisiert. Sie besteht generell aus einer hydrophoben Region von etwa 10 bis 20 Aminosäuren, die als "core" bezeichnet wird, an deren Aminoterminus eine kurze Peptidsequenz
15 ("pre-core") gebunden ist, die meist eine positiv geladene Aminosäure (oder mehrere) aufweist. Zwischen dem Carboxy-terminus der hydrophoben Region und dem Aminoterminus des "reifen" transportierten Proteins liegt eine kurze Peptidsequenz ("post-core"), welche die Spaltstelle
20 ("splice-site") enthält und für eine günstige räumliche Anordnung sorgt.

- Aus der US-Patentschrift 4 411 994 ist bekannt, das Gen für ein zu exprimierendes Protein an ein bakterielles Gen
25 zu koppeln, das für ein extracelluläres oder periplasmatisches Trägerprotein codiert, um so den Transport des gewünschten Proteins aus dem Cytoplasma zu bewerkstelligen. Für dieses Verfahren muß man ein wirtseigenes, bakterielles Gen für ein periplasmatisches, "outer membrane"- oder ein
30 extracelluläres Protein isolieren. Dieses Gen wird dann mit einem Restriktionsenzym geschnitten, das Gen für das zu transportierende Protein in die gebildete Schnittstelle inseriert und die Wirtszelle mit einem Vektor transformiert, der das so gebildete Fusionsgen enthält. Die Isolierung
35 des natürlichen Gens und seine Charakterisierung zur Auswahl geeigneter Schnittstellen ist außerordentlich aufwendig.

Erfindungsgemäß wird dieser Aufwand dadurch vermieden, daß von einer synthetischen Signalsequenz Gebrauch gemacht wird.

- 5 Die Erfindung betrifft deshalb eine synthetische Signalsequenz zum Transport von Proteinen in Expressionssystemen, die dadurch gekennzeichnet ist, daß die DNA im wesentlichen einer natürlichen Signalsequenz entspricht, aber eine oder mehrere Schnittstellen für Endonucleasen aufweist, die in
- 10 der natürlichen DNA nicht enthalten sind. Weitere Aspekte der Erfindung und bevorzugte Ausgestaltungen sind im folgenden wiedergegeben bzw. in den Patentansprüchen niedergelegt.
- 15 Die DNA soll "im wesentlichen" der einer natürlichen Signalsequenz entsprechen. Hierunter ist zu verstehen, daß das exprimierte Signalpeptid mit dem natürlichen weitgehend oder völlig übereinstimmt, im letzteren Falle also lediglich auf der DNA-Ebene insofern ein Unterschied besteht,
- 20 als die synthetische DNA mindestens eine Schnittstelle aufweist, die in der natürlichen DNA-Sequenz nicht enthalten ist. Dieser Einbau der erfindungsgemäßen Schnittstelle bedingt also eine mehr oder weniger große Abweichung von der natürlichen Sequenz, wobei unter Umständen auf Codons zu-
- 25 rückgegriffen werden muß, die bekanntermaßen vom jeweiligen Wirtsorganismus weniger bevorzugt werden. Überraschenderweise ist hiermit jedoch kein Nachteil in der Expression verknüpft. Das gezielte "Maßschneidern" des synthetischen Gens bringt vielmehr so viele Vorteile mit sich, daß ein
- 30 eventueller Nachteil durch den "unnatürlichen" Codon-Gebrauch im allgemeinen weit überkompensiert wird. Es hat sich sogar gezeigt, daß der Ersatz des Startcodons GTG, das im Gen der Alkalischen Phosphatase von E. coli vorkommt, durch ATG zu einer stark erhöhten Expression führt. Ein
- 35 besonderer Vorteil der Erfindung liegt darin, daß die Wirtszelle weniger Ballast-Protein produzieren muß, da das zu exprimierende Gen direkt an das 3'-Ende der synthetischen DNA-Signalsequenz angeküpft werden kann. Darüber hinaus ergeben sich insofern Vorteile, als sich bei der

Konstruktion der synthetischen DNA an den Enden überstehende DNA-Sequenzen für bestimmte Restriktionserkennungsstellen vorsehen lassen, die eine Klonierung dieser Sequenz erlauben und im Falle unterschiedlicher Erkennungsstellen einen definierten Einbau in einen Klonierungsvektor gestatten. Hierdurch wird nach dem "Baukastenprinzip" ein Einbau in beliebige Vektoren ermöglicht.

Interne Erkennungsstellen für Restriktionsenzyme gestatten die Ankoppelung beliebiger homologer oder heterologer Gene im richtigen Leserahmen. Über diese internen Schnittstellen können auch auf einfache Weise Variationen in die DNA der Signalsequenzen eingeführt werden, die zu in der Natur nicht vorkommenden Praesequenzen führen.

Zweckmäßig werden diese internen Schnittstellen in die Bereiche vor und nach der hydrophoben Region gelegt, insbesondere in die "post-core"-Region, wobei die Spaltstelle und/oder deren Umgebung modifiziert werden kann. Es ist natürlich auch möglich, die "core"-Region in an sich bekannter Weise zu modifizieren.

Unter Berücksichtigung bekannter Regeln (G. von Heijne, J. Mol. Biol. 173 (1984) 243 - 251) kann über geeignete Schnittstellen im Genabschnitt, der für den carboxyterminalen Teil des Praepeptids codiert, die Signalpeptidase-Schnittstelle derart geplant werden, daß kein Fusionsprotein, sondern direkt das gewünschte, i.a. eukaryotische Peptid in nativer Form exprimiert wird. Gene natürlichen Ursprungs lassen eine solche Prozessierung im allgemeinen nicht zu.

Als Signalsequenzen kommen im Prinzip alle literaturbekannten Signalsequenzen (M.E.E. Watson; Nucleic Acids Res. 12 (1984), 5145 - 5164), Variationen derselben sowie daraus abgeleitete "idealisierte" Signalsequenzen (D. Perlman und H.O. Halvorson; J. Mol. Biol. 167 (1983), 391 - 409) in Betracht.

Bevorzugte Wirtsorganismen sind E. coli, Staphylococcus aureus, Bacillus subtilis, Streptomyces, Bacillus amyloliquifaciens, Bacillus cereus, Bacillus licheniformis, Pseudomonas, Saccharomyces, Spodoptera frugiperda und Zell-
 5 linien höherer Organismen wie pflanzliche oder tierische Zellen.

Prinzipiell lassen sich alle diejenigen Proteine prokaryotischen oder eukaryotischen Ursprungs durch Transportexpression
 10 gewinnen, die membrangängig sind. Bevorzugt sind aber pharmazeutisch bedeutende Peptidprodukte wie Hormone, Lymphokine, Interferone, Blutgerinnungsfaktoren und Vakzine, die in der Natur auch als Peptide mit einer aminoterminalen Praesequenz codiert sind. Diese eukaryotische Praesequenz wird
 15 jedoch in der Regel in den prokaryotischen Wirtsorganismen nicht von den wirtseigenen Signalpeptidasen abgespalten.

In E. coli sind die Gene für die periplasmatischen und "outer membrane"-Proteine zur Transportexpression geeignet,
 20 wobei die erstgenannten das Produkt ins Periplasma, die letztgenannten aber eher an die äußere Membran dirigieren.

Als Beispiel wird die DNA-Signalsequenz des periplasmatischen Proteins Alkalische Phosphatase, das in E. coli
 25 hochexprimierbar ist, angeführt, ohne daß die Erfindung darauf beschränkt wäre.

Im folgenden ist die Prae-Sequenz einschließlich der ersten zwanzig Aminosäuren der Alkalischen Phosphatase von
 30 E. coli dargestellt:

1	5	10
Met-Lys-Gln-Ser-Thr-Ile-Ala-Leu-Ala-Leu-Leu-Pro-Leu-Leu-		
15	20	25
Phe-Thr-Pro-Val-Thr-Lys-Ala-Arg-Thr-Pro-Glu-Met-Pro-Val-		
35	30	35
Leu-Glu-Asn-Arg-Ala-Ala-Gln-Gly-Asn-Ile-Thr-Ala-Pro		40

†= bevorzugte Spaltstelle der Signalpeptidase:

Es hat sich gezeigt, daß bis zu etwa 40, meist etwa 20 zusätzliche Aminosäuren des reifen Proteins für eine korrekte Prozessierung ausreichend sind. In vielen Fällen genügen jedoch auch weniger zusätzliche Aminosäuren, beispielsweise etwa 10, vorteilhaft etwa 5. Da eine kürzere Proteinkette

5 eine geringere Beanspruchung des Protein-Biosyntheseapparats der Wirtszelle bedingt, ist eine vorteilhafte Ausführungsform der Erfindung in der DNA-Sequenz I (Anhang) niedergelegt, die für die Praesequenz der Alkalischen Phosphatase sowie zusätzlicher 5 Aminosäuren des reifen

10 Proteins codiert. Die DNA-Sequenz I entspricht bis auf einige Triplet-Variationen - nämlich solchen, die singuläre Restriktionsenzym-Schnittstellen einführen und das Startcodon GTG durch ATG ersetzen - der natürlichen Sequenz der Alkalischen Phosphatase. An den Enden des codierenden

15 Stranges befinden sich "überhängende" DNA-Sequenzen entsprechend der Restriktionsendonuclease EcoR I, die den Einbau in gängige Klonierungsvektoren, beispielsweise die handelsüblichen Plasmide wie pBR 322, pUC 8 oder pUC 12 erlauben. Zusätzlich wurden innerhalb des Gens der DNA-

20 Sequenz I eine Reihe von weiteren singulären Schnittstellen für Restriktionsenzyme eingebaut, die einerseits die Ankopplung heterologer Gene an der richtigen Stelle und im gewünschten Leserahmen ermöglichen und andererseits auch die Durchführung von Variationen erlauben:

25

Restriktionsenzym	Schnitt nach Nucleotid-Nr.
	(im codierenden Strang)

	Sau 3 A	19
30	Pvu I	22
	Hpa II	54) (im natürlichen
	Nci I	54) Gen vorhanden)
	Alu I	66
	Hph I	68
35	Ava II	70

Es ist selbstverständlich auch möglich, die überhängenden Sequenzen so zu konstruieren, daß sie verschiedenen Restriktionsenzymen entsprechen, was dann den Einbau in geeignete Vektoren in definierter Orientierung erlaubt. Der Fachmann
5 wird hierbei abwägen, ob der mit dem Konstruieren des Gens und seinem gezielten Einbau verbundene Aufwand schwerer wiegt als die zusätzliche Selektionsarbeit, die mit einem Einbau in beiden Orientierungsrichtungen bei identischen überstehenden Enden verbunden ist.

10

Die DNA-Sequenz I läßt sich aus 6 Oligonucleotiden mit einer Länge von 26 - 31 Basen aufbauen, indem diese zunächst chemisch synthetisiert und dann über "sticky ends" von 6 Nucleotiden enzymatisch verknüpft werden. Der Einbau des
15 synthetischen Gens in Klonierungsvektoren, beispielsweise in die genannten handelsüblichen Plasmide, erfolgt in an sich bekannter Weise.

Als Beispiel für die Expression eines eukaryotischen Gens
20 in E. coli mit Hilfe einer erfindungsgemäßen Präesequenz wird im folgenden die Synthese von Affen-Proinsulin beschrieben: Man konstruiert eine DNA-Sequenz, bei der hinter einer chemisch-synthetischen Regulationsregion, bestehend aus einem bakteriellen Promotor, einem lac-Operator und
25 einer ribosomalen Bindungsstelle (Deutsche Patentanmeldung P 34 30 683.8), 6 - 14 Nucleotide von der Ribosomen-Bindungsstelle entfernt, an eine anschließende Erkennungssequenz für EcoR I die DNA-Sequenz I und daran anschließend
30 angeordnet ist. Das exprimierte Proinsulin-Fusionspeptid enthält am Aminotermius noch 9 zusätzliche Aminosäuren, die enzymatisch oder chemisch abgespalten werden können.

Der Einbau des synthetischen Gens in pUC 8 sowie die Konstruktion von Expressionsplasmiden, die eukaryotische Gene
35 an die DNA-Sequenz I gekoppelt enthalten, erfolgt in an

sich bekannter Weise. Hierzu kann auf das Lehrbuch von Maniatis (Molecular Cloning, Maniatis et al., Cold Spring Harbor, 1982) verwiesen werden. Die Transformation der so erhaltenen Hybridplasmide in geeignete Wirtsorganismen, vor-
5 teilhaft E. coli, ist ebenfalls an sich bekannt und in dem vorstehend genannten Lehrbuch eingehend beschrieben. Die Gewinnung der exprimierten Proteine und deren Reinigung ist ebenfalls beschrieben.

- 10 In den folgenden Beispielen werden noch einige Ausgestaltungen der Erfindung im einzelnen erläutert, woraus sich die Vielzahl der möglichen Abwandlungen (und Kombinationen) für den Fachmann ergeben. Prozentangaben beziehen sich hierbei auf das Gewicht, wenn nichts anderes angegeben ist.

15

Beispiele

1. Chemische Synthese eines einzelsträngigen Oligonucleotids

- 20 Am Beispiel des Genbausteins Ia, der die Nucleotide 1 - 29 des codierenden Strangs umfaßt, wird die Synthese der Genbausteine erläutert. Nach bekannten Methoden (M.J. Gait et al., Nucleic Acids Res. 8 (1980) 1081 - 1096)) wird das am 3'-Ende stehende Nucleosid, im vorliegenden Falle
25 also Guanodin (Nucleotid Nr. 29), an Kieselgel (FRACTOSIL, Firma Merck) über die 3'-Hydroxyfunktion covalent gebunden. Hierzu wird zunächst das Kieselgel unter Abspaltung von Ethanol mit 3-(Triethoxysilyl)-propylamin umgesetzt, wobei eine Si-O-Si-Bindung entsteht. Das Guanodin
30 wird als N^{2'}-Isobutyryl-3'-O-succinoyl-5'-dimethoxytritylether in Gegenwart von Paranitrophenol und N,N'-Dicyclohexylcarbodiimid mit dem modifizierten Träger umgesetzt, wobei die freie Carboxygruppe der Succinoylgruppe den Aminorest der Propylaminogruppe acyliert.

35

In den folgenden Syntheseschritten wird die Basenkomponente als 5'-O-Dimethoxytrityl-nucleosid-3'-phosphorigsäure-monomethylester-dialkylamid oder -chlorid eingesetzt, wobei das Adenin als N⁶-Benzoyl-Verbindung, das Cytosin als N⁴-Benzoyl-Verbindung, das Guanin als N²-Isobutyryl-Verbindung und das keine Aminogruppe enthaltende Thymin ohne Schutzgruppe vorliegen.

50 mg des polymeren Trägers, der 2 µmol Guanosin gebunden enthält, werden nacheinander mit den folgenden Agentien behandelt:

- a) Nitromethan
- b) gesättigte Zinkbromidlösung in Nitromethan mit 1 % Wasser
- c) Methanol
- d) Tetrahydrofuran
- e) Acetonitril
- f) 40 µmol des entsprechenden Nucleosidphosphits und 200 µmol Tetrazol in 0,5 ml wasserfreiem Acetonitril (5 Minuten)
- g) 20 % Acetanhydrid in Tetrahydrofuran mit 40 % Lutidin und 10 % Dimethylaminopyridin (2 Minuten)
- h) Tetrahydrofuran
- i) Tetrahydrofuran mit 20 % Wasser und 40 % Lutidin
- j) 3 % Jod in Kollidin/Waasser/Tetrahydrofuran im Volumenverhältnis 5 : 4 : 1
- k) Tetrahydrofuran und
- l) Methanol.

30

Unter "Phosphit" wird hierbei der Desoxyribose-3'-monophosphorigsäure-monomethylester verstanden, wobei die dritte Valenz durch Chlor oder eine tertiäre Aminogruppe, beispielsweise einen Morpholinorest, abgesättigt ist.

35

Die Ausbeuten der einzelnen Syntheseschritte können je-

weils nach der Detritylierungsreaktion (b) spektrophotometrisch durch Messung der Absorption des Dimethoxytritylkations bei der Wellenlänge von 496 nm bestimmt werden.

- 5 Nach abgeschlossener Synthese des Oligonucleotids werden die Methylphosphatschutzgruppen des Oligomers mit Hilfe von p-Thiokresol und Triethylamin abgespalten. Anschließend wird durch 3-stündige Behandlung mit Ammoniak das Oligonucleotid vom festen Träger abgetrennt. Eine
10 2- bis 3-tägige Behandlung der Oligomeren mit konzentriertem Ammoniak spaltet die Aminoschutzgruppen der Basen quantitativ ab. Das so erhaltene Rohprodukt wird durch Hochdruckflüssigkeitschromatographie (HPLC) oder durch Polyacrylamid-Gelelektrophorese gereinigt.

15

Ganz entsprechend werden auch die übrigen Genbausteine Ib - If synthetisiert, deren Nucleotidfolge aus der DNA-Sequenz II (Anhang) hervorgeht.

- 20 2. Enzymatische Verknüpfung der einzelsträngigen Oligonucleotide zur DNA-Sequenz I

- Die endständigen Oligonucleotide Ia und If werden nicht phosphoryliert. Damit wird eine Oligomerisation über die
25 überhängenden Enden vermieden. Zur Phosphorylierung der Oligonucleotide Ib, Ic, Id und Ie werden jeweils 1 nmol dieser Verbindungen mit 5 nmol Adenosintriphosphat und vier Einheiten T4-Polynucleotid-Kinase in 20 µl 50 mM Tris-HCl-Puffer (pH 7,6), 10 mM Magnesiumchlorid und
30 10 mM Dithiothreitol (DTT) 30 Minuten bei 37°C behandelt. Das Enzym wird durch fünfminütiges Erhitzen auf 95°C deaktiviert. Anschließend werden die Oligonucleotide Ia bis If vereinigt und zum Doppelstrang hybridisiert, indem man sie in einer 20 mM KCl-Lösung erhitzt und dann langsam (im Laufe von 2 Stunden) auf 16°C ab-
35 kühlt. Die Ligation zum DNA-Fragment gemäß DNA-Sequenz I

erfolgt durch Reaktion in 40 µl 50 mM Tris-HCl-Puffer (20 mM Magnesiumchlorid und 10 mM DTT) mit Hilfe von 100 Einheiten T4-DNA-Ligase bei 15°C im Laufe von 18 Stunden.

5

Die Reinigung des Genfragments erfolgt durch Gelelektrophorese auf einem 10 %igen Polyacrylamidgel (ohne Harzstoffzusatz, 20 x 40 cm, 1 mm Dicke), wobei als Markersubstanz ϕ X 174 DNA (Fa. BRL), geschnitten mit Hinf I, oder pBR 322, geschnitten mit Hae III, dient:

10

3. Einbau des Genfragmentes in pUC 8

Das handelsübliche Plasmid pUC 8 wird in bekannter Weise mit der Restriktionsendonuclease EcoR I nach den Angaben des Herstellers geöffnet. Der Verdauungsansatz wird auf einem 5 %igen Polyacrylamidgel durch Elektrophorese in bekannter Weise aufgetrennt und die DNA durch Anfärben mit Ethidiumbromid oder durch radioaktive Markierung ("Nick-Translation" nach Maniatis, a.a.O.) sichtbar gemacht. Die Plasmidbande wird anschließend aus dem Acrylamidgel ausgeschnitten und elektrophoretisch vom Polyacrylamid abgetrennt.

20

25 4. Einbau der DNA-Sequenz I in ein Expressionsplasmid

Das Expressionsplasmid pWI 6 mit der Information für das Affen-Proinsulin wird wie folgt konstruiert:

30

10 µg des Plasmids pBR 322 werden mit den Restriktionsendonucleasen Eco RI/Pvu II verdaut und anschließend durch eine "fill-in"-Reaktion mit Klenow-Polymerase an der Eco RI-Schnittstelle aufgefüllt. Nach gelelektrophoretischer Auftrennung in einem 5 %igen Polyacrylamidgel läßt sich das Plasmidfragment von 2293 Bp Länge durch Elektroelution gewinnen (Figur 1).

35

- Aus dem Plasmid pBR 322 mit integrierter Affenpräproinsulin-DNA (Wetekam et al., Gene 19 (1982) 179 - 183) wird diese durch Verdauung mit den Restriktionsendonucleasen Hind III und Mst I isoliert und in das Plasmid
- 5 pUC 9 wie folgt rekloniert: Das Plasmid pUC 9 wird mit dem Enzym Bam HI gespalten, die Spaltstelle in einer Standard-"fill-in"-Reaktion mit Klenow-Polymerase ("large fragment") aufgefüllt, mit dem Restriktionsenzym Hind III nachgeschnitten und die DNA gelelektrophoretisch in einem 5 %igen Polyacrylamidgel von den übrigen
- 10 DNA-Fragmenten abgetrennt. In das geöffnete Plasmid wird das isolierte Insulin-DNA-Fragment von etwa 1250 Bp Länge integriert. Zur Abtrennung der "untranslated region" und der Präsequenz wird mit Hae III verdaut und
- 15 das 143 Bp lange Fragment zur Abspaltung der letzten beiden Nucleotide aus der Präsequenz unter limitierenden Enzymbedingungen mit Bal 31 verdaut. Man erhält so am Amino-Terminus als erstes Codon TTT, das für Phenylalanin als erste Aminosäure der B-Kette steht.
- 20 An dieses Fragment wird nun ein für Eco RI spezifischer Adaptor in einer "blunt-end" Ligationsreaktion ligiert:
- a) 5' AAT TAT GAA TTC GCA ATG
25 Eco RI TA CTT AAG CGT TAC
- b) 5' AAT TAT GAA TTC GCA AGA
 Eco RI TA CTT AAG CGT TCT
- 30 Um eine Polymerisierung der Adaptern zu vermeiden, wurden diese unphosphoryliert in die Ligationsreaktion eingesetzt. Der Adaptor a) weist am Ende ein Codon für Methionin auf, der Adaptor b) das Codon für Arginin. Nach der Variante a) wird somit ein Genprodukt erhalten, bei
- 35 dem durch Bromcyanspaltung eine Abtrennung des bakteriellen Anteils möglich ist, während die Variante b) eine Trypsinspaltung erlaubt.

Das Ligationsprodukt wird mit Mbo II verdaut, Nach gelelektrophoretischer Auftrennung erhält man ein DNA-Fragment von 79 Bp Länge mit der Information für die Aminosäuren Nr. 1 bis 21 der B-Kette.

- 5 Das Gen für die restliche Information des Proinsulinmoleküls (einschließlich einer G-C-Sequenz aus der Klonierung und 21 Bp aus dem pBR 322 im Anschluß an das Stop-Codon) erhält man aus dem pUC 9-Plasmid mit der
- 10 kompletten Präproaffensulin-Information durch Verdauung mit Mbo II/Sma I und Isolierung eines DNA-Fragments von etwa 240 Bp Länge. Durch Ligieren der beiden Proinsulin-Fragmente erhält man das korrekte Ligationsprodukt von etwa 320 Bp Länge (inklusive des
- 15 Adaptors von 18 Bp). Dieses so konstruierte Proinsulin-DNA-Fragment kann nun mit einer Regulationsregion über die Eco RI-negative Schnittstelle zusammenligiert werden.
- 20 Die gesamte Reaktionsfolge zeigt die Fig. 2, in der A, B und C die DNA für die jeweiligen Peptidketten des Proinsulinmoleküls, Ad den (dephosphorylierten) Adaptor (a oder b) und P die DNA für die Präsequenz des Affenpräproinsulins bedeuten.
- 25 Eine chemisch-synthetische Regulationsregion bestehend aus einer Erkennungssequenz für Bam HI, dem lac-Operator, einem bakteriellen Promotor und einer ribosomalen Bindungsstelle (RB) mit einem ATG-Start-Codon, 6
- 30 bis 14 Nucleotide von der RB entfernt, und einer anschließenden Erkennungssequenz für Eco RI (Figur 3) wird über die gemeinsame Eco RI-Überlappingsregion mit dem nach dem vorstehenden Beispiel erhaltenen Proinsulin-Fragment ligiert. Nach einer Doppelverdauung mit Sma
- 35 I/Bam HI und einer "fill-in" Reaktion der Bam HI Schnittstelle mit dem Kleenow-Fragment wird das Ligationsprodukt (ungefähr 420 Bp) gelelektrophoretisch isoliert.

Das so erhaltene Fragment läßt sich anschließend über eine "blunt-end" Ligierung in das pBR 322-Teilplasmid gemäß Figur 1 hineinligieren (Figur 4). Man erhält das Hybridplasmid pWI 6.

5

Nach Transformation in den E. coli-Stamm HB 101 und Selektion auf Ampicillin-Platten wurde die Plasmid-DNA individueller Klone auf die Integration eines 420

10

Bp-Fragments mit der Regulationsregion und dem Bal 31-verkürzten Proinsulingen getestet. Zum Nachweis der korrekten Verkürzung des Proinsulingens durch Bal 31 (Figur 2) wurden die Plasmide mit dem integrierten Proinsulinfragment von der Eco RI-Schnittstelle aus sequenziert. Von 60 sequenzierten Klonen hatten drei die gewünschte Verkürzung von zwei Nucleotiden (Figur 4).

15

1 µg des Plasmides pWI 6 wird mit dem Restriktionsenzym Eco RI geschnitten und anschließend in Gegenwart von 30 ng DNA-Sequenz I bei 16°C in 6 Stunden zusammenligiert. Nach Transformation in E. coli HB 101 werden aus individuellen Klonen Plasmide isoliert und mit Hilfe einer Restriktionsenzymanalyse auf die Integration der DNA-Sequenz I getestet. 7 % aller Klone enthielten das Plasmid pWI 6 mit der DNA-Sequenz I integriert.

25

Die Richtung dieser Integrationsreaktion läßt sich mit Standardmethoden der Restriktionsenzymanalyse über eine Doppelverdauung mit Hind III/Pvu I eindeutig ermitteln. Das Plasmid pWI 6 mit einer DNA-Sequenz I in korrekter Leserichtung zum Proinsulin integriert ist als pWIP 1 in Figur 5 dargestellt.

30

Dieses Plasmid kann anschließend in verschiedene E. coli-Stämme hineintransformiert werden, um die Synthesekapazität der einzelnen Stämme auszutesten.

35

Die Expression der Praesequenz-Proinsulin-Fusion in *E. coli* wird folgendermaßen bestimmt:

5 1 ml einer mit IPTG (Isopropyl-p-D-thiogalaktopyranosid) induzierten Bakterienkultur bei einer optischen Dichte von 1,0 und einer Induktionsdauer von 1 Stunde werden mit PMSF (Phenylmethylsulfonylfluorid) in einer Endkonzentration von $5 \times 10^{-4} \text{ M}$ gestoppt, in Eis gekühlt und abzentrifugiert. Der Zelniederschlag wird anschließend
10 in 1 ml Puffer (10 mM Tris-HCl, pH 7,6; 40 mM NaCl) gewaschen, abzentrifugiert und in 200 µl Puffer (20 % Saccharose; 20 mM Tris-HCl, pH 8,0; 2 mM EDTA) resuspendiert, 10 Minuten bei Zimmertemperatur inkubiert, abzentrifugiert und sofort in 500 µl H_2O bidest. resuspendiert. Nach 10 Minuten Inkubationszeit in Eis werden die
15 "Schock-lysierten" Bakterien abzentrifugiert und der Überstand eingefroren. Dieser Überstand wird in einem Standard-Insulin-RIA (Amersham) auf Proinsulingehalt getestet.

20 Der Bakterienniederschlag wird noch einmal in 200 µl Lysozypuffer (20 % Saccharose; 2 mg/ml Lysozym; 20 mM Tris-HCl, pH 8,0; 2 mM EDTA) resuspendiert, 30 Minuten in Eis inkubiert, 3 mal 10 Sekunden mit Ultraschall be-
25 handelt und anschließend abzentrifugiert. Der so resultierende Überstand wird als Plasmafraktion auf Gehalt an Proinsulin in einem Radioimmunoassay getestet.

30 Individuelle Bakterienklone, welche das Plasmid pWIP 1 enthalten, wurden auf ihre Synthesekapazität und ihre Transportfähigkeit des Proinsulin-Praesequenz-Produktes untersucht. Es konnte gezeigt werden, daß sämtliche Bakterienklone in erwarteter Weise das produzierte Proinsulin zu ungefähr 90 % in den periplasmatischen Raum transportierten. Ungefähr 10 % der Proinsulin-RIA-Aktivität waren noch in der Plasmafraktion zu finden.
35

DNA-Sequenz I

Triplet Nr.			1	2	3	
Aminosäure Nr.			Met	Lys	Gln	
Nucleotid Nr.			5	10		
Codierender Strang	5'	AA	TTC	ATG	AAA	CAA
nichtcod. Strang	3'		G	TAC	TTT	GTT

4	5	6	7	8	9	10	11	12	13
Ser	Thr	Ile	Ala	Leu	Ala	Leu	Leu	Pro	Leu

15	20		25		30	35		40	
AGC	ACG	ATC	GGA	CTG	GCA	CTC	TTA	CCG	TTA
TCG	TGC	TAG	CGT	GAC	CGT	GAG	AAT	GGC	AAT

14	15	16	17	18	19	20	21	22	23
Leu	Phe	Thr	Pro	Val	Thr	Lys	Ala	Arg	Thr

45	50		55		60	65		70	
CTG	TTT	ACC	CCG	GTG	ACA	AAA	GCT	CGG	ACC
GAC	AAA	TGG	GGC	CAC	TGT	TTT	CGA	GCC	TGG

24	25	26
Pro	Glu	Met

75	80		84		
CCA	GAA	ATG	G		3'
GGT	CTT	TAC	CTT	AA	5'

18.
- 16 -

DNA-Sequenz II:

← Ia →
5' AA TTC ATG AAA CAA AGC ACG ATC GCA CTG
3' G TAC TTT GTT TCG TGC TAG CGT GAC
Eco RI ← Ib →

← Ic →
GCA CTC TTA CCG TTA CTG TTT ACC CCG
CGT GAG AAT GGC AAT GAC AAA TGG GGC
→ Id →

← Ie → Eco RI
GTG ACA AAA GCT CGG ACC CCA GAA ATG G
CAC TGT TTT CGA GCC TGG GGT CTT TAC CTT AA
→ If →

Nummer:
 Int. Cl.4:
 Anmeldetag:
 Offenlegungstag:

34 36 818
 C 12 N 15/00
 6. Oktober 1984
 10. April 1986

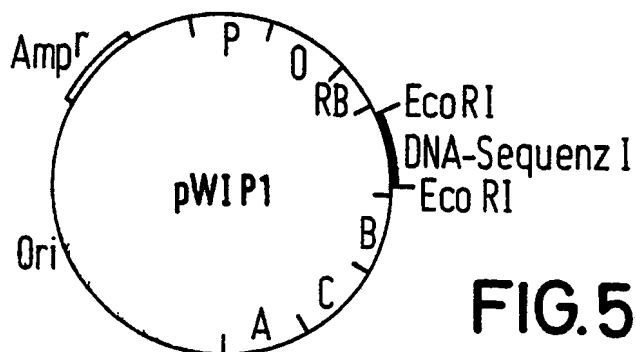
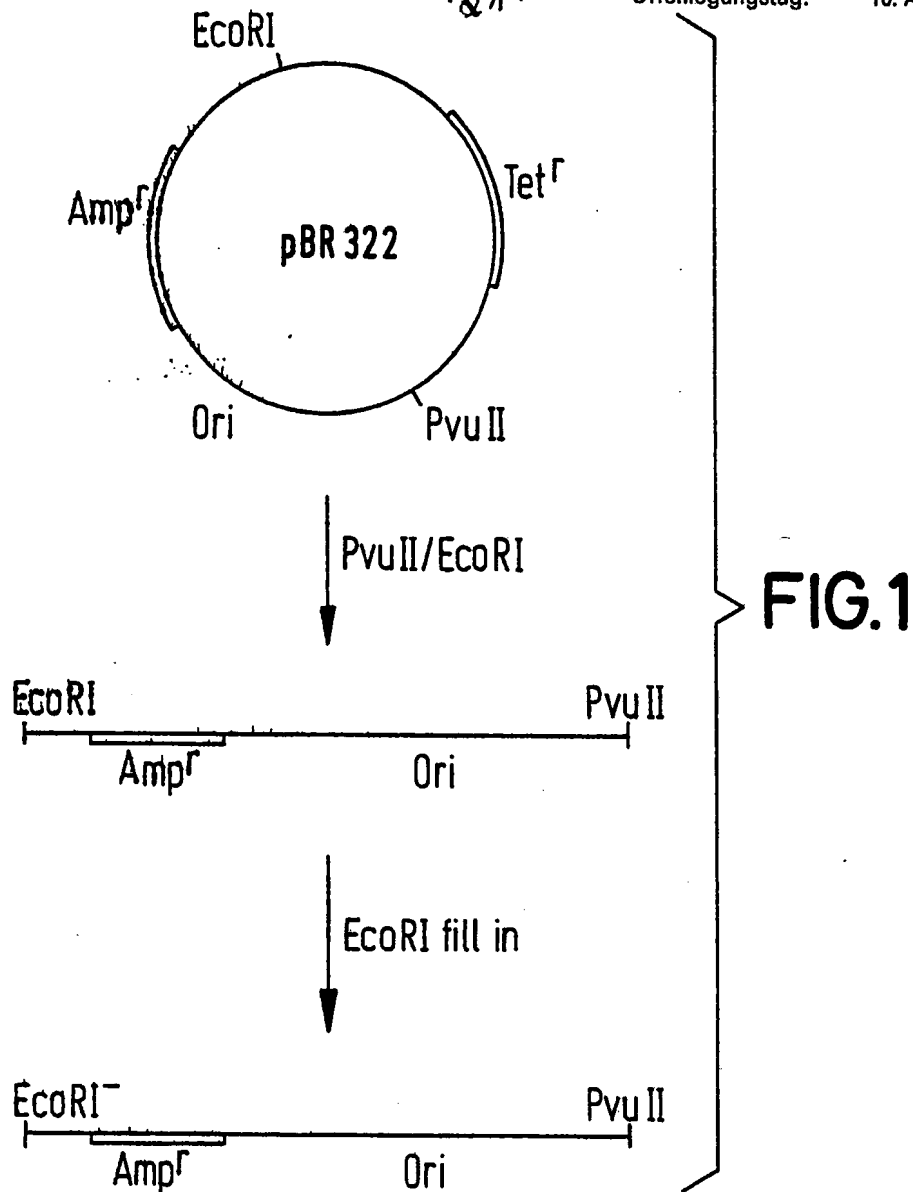
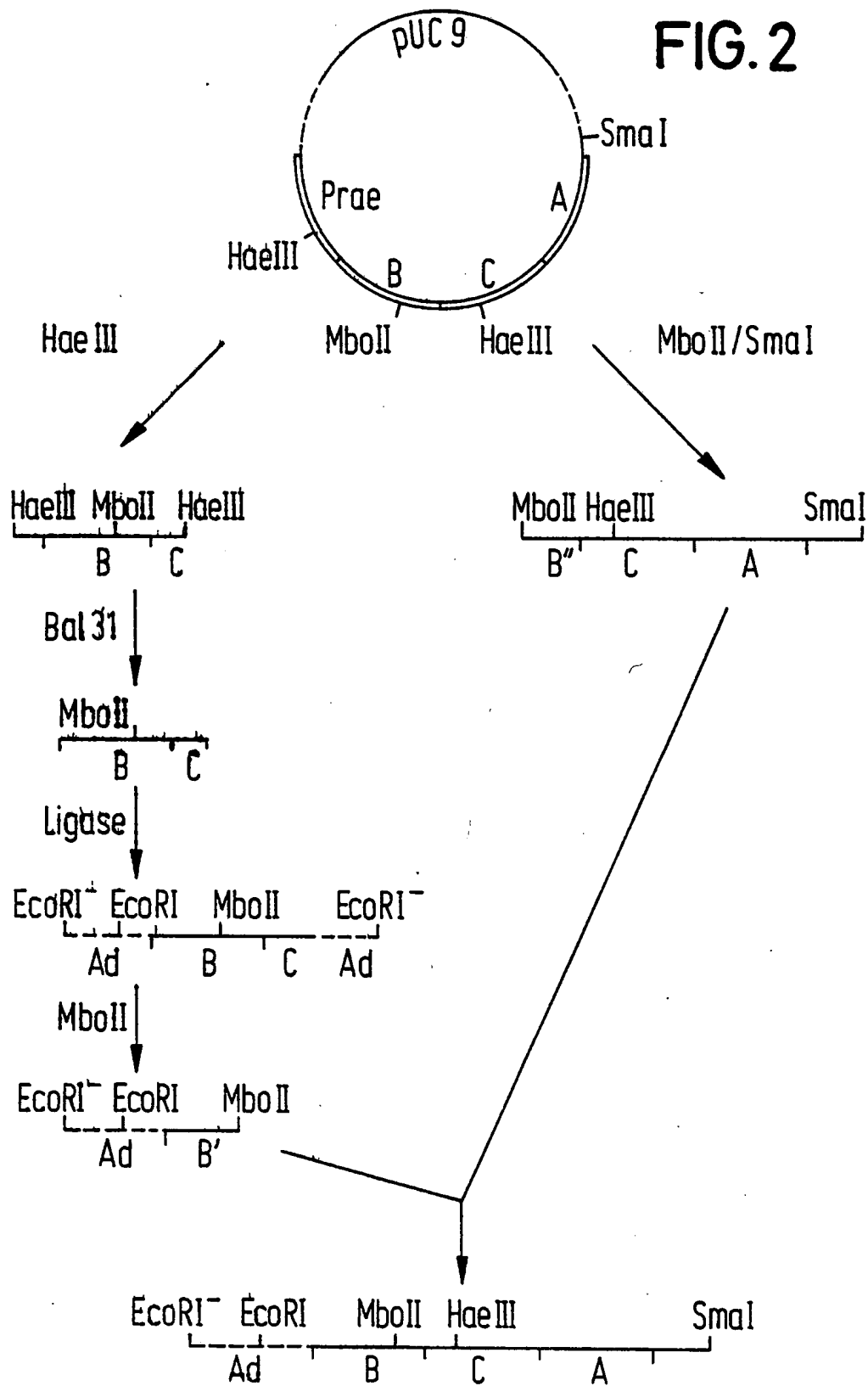


FIG. 2



20.

FIG. 3

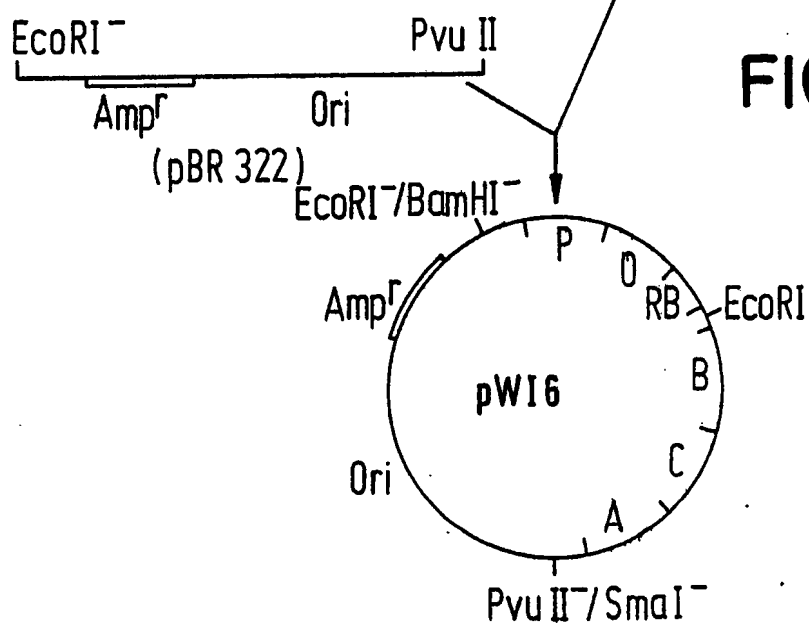
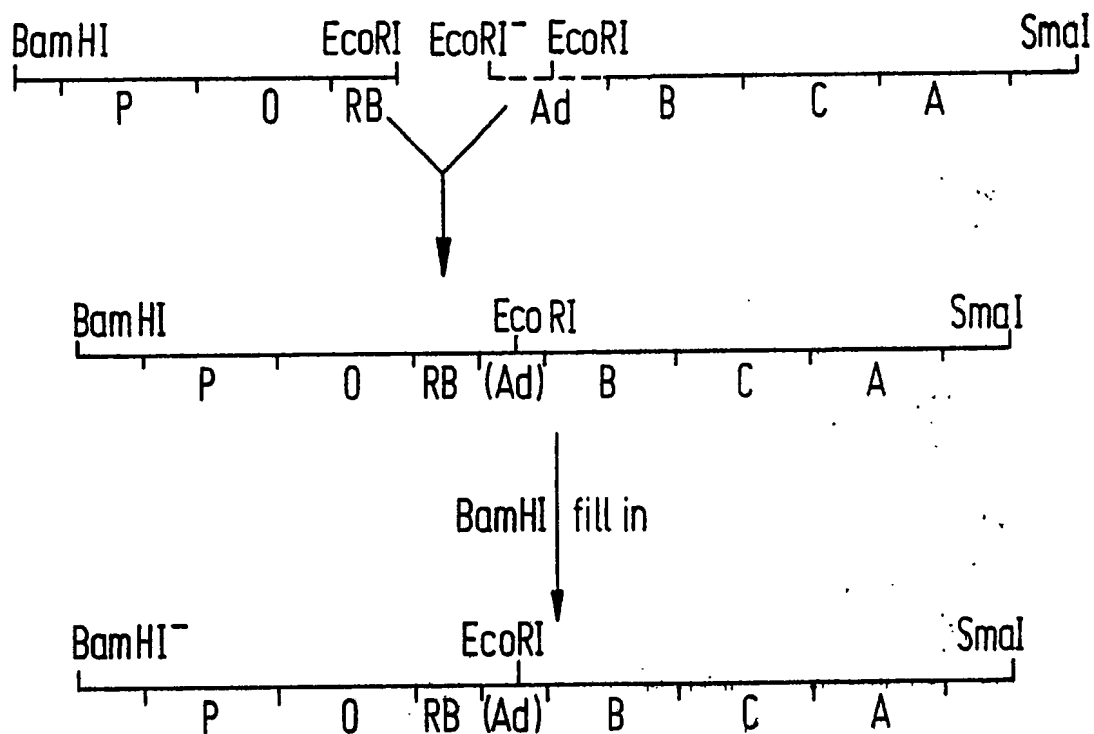


FIG. 4